

Capítulo 7

MARC 21 e XML como ferramentas para a consolidação da catalogação cooperativa automatizada Uma revisão de literatura

Plácida Leopoldina Ventura Amorim da Costa Santos
Adriana Nascimento Flamino

Introdução

A Catalogação Cooperativa Automatizada, que hoje nos possibilita a cooperação e o compartilhamento de informações catalográficas e bibliográficas entre diversas instituições no ambiente automatizado, somente se tornou possível com a criação e adoção do Machine Readable Cataloging – MARC, um Formato de Intercâmbio desenvolvido e testado pela Library of Congress – LC (Biblioteca do Congresso Norte Americano), na década de 1960.

Devido aos avanços das tecnologias em informação, do crescente uso da Internet e dos bancos de dados, houve a necessidade da criação de ferramentas que otimizassem cada vez mais as atividades de organização, recuperação e intercâmbio de informações. Segundo Marcondes (2001, p. 61), “de nada adianta a informação existir, se quem dela necessita não sabe da sua existência, ou se ela não puder ser encontrada”.

Carvalho já em 1990 alertava:

A ciência da atualidade está produzindo uma quantidade de dados que vai atingir as proporções de um dilúvio, milhões de vezes superior à capacidade humana de manipulá-los ou estudá-los. Os pesquisadores produziram e armazenaram em máquina, nos últimos 20 anos, um volume de informações equivalente a tudo que foi descoberto nos 20 mil anos de história da civilização do homem (CARVALHO, 1990, p. 24).

Dessa forma, novos formatos de intercâmbio e novas linguagens surgem. A eXtensible Markup Language – XML é um desses desenvolvimentos que tem como propósito facilitar o gerenciamento, armazenamento e transmissão de dados, via Internet.

Portanto, para uma Unidade de Informação tornar-se útil é primordial que ela esteja conectada ao mundo globalizado, para assim recuperar, armazenar, representar e disseminar as informações atualizadas e necessárias ao usuário, independente de seu suporte informacional. Sendo assim, as ferramentas das novas tecnologias são as portas de entrada para essa realidade.

A revisão de literatura é uma das metodologias de que os cientistas necessitam e se utilizam hoje para identificar, conhecer e acompanhar o desenvolvimento de pesquisas em suas áreas de atuação, pois a gama de informações existente hoje e a sua dispersão física dificultam cada vez mais este acompanhamento, tanto pela impossibilidade de acessar tudo o que existe e controlar toda a bibliografia, como também pela limitação humana de absorver toda a informação relevante e disponível. As autoras Noronha e Ferreira (2000, p.197) comentam com muita pertinência:

[...] os trabalhos de revisão, contribuem para o desenvolvimento do conhecimento científico, principalmente porque comparam informação de fontes diferentes; compactam o conhecimento existente; identificam especializações emergentes; direcionam pesquisas para novas áreas; notificam os pesquisadores periodicamente sobre a literatura publicada em dada época; promovem um serviço de alerta para campos correlatos; dão suporte à busca bibliográfica, constituindo-se em fonte inicial básica para a elaboração de projetos de pesquisa; auxiliam indiretamente o ensino, dando suporte a trabalhos acadêmicos e, finalmente, oferecem um feedback através da avaliação do trabalho publicado.

Pode-se perceber a necessidade de estudos como esse na área da biblioteconomia, inclusive para que o profissional bibliotecário tenha mais conhecimento, domínio e utilize as atuais ferramentas de trabalho, os computadores, formatos de intercâmbio bibliográfico e catalográfico, linguagens de marcação, protocolos e outros proce-

dimentos envolvidos no processo da Catalogação Cooperativa Automatizada, do intercâmbio e do compartilhamento de dados bibliográficos e catalográficos.

Portanto, este é um tema atual e necessário para a área da biblioteconomia no que diz respeito à avaliação das formas de representação para o tratamento adequado das informações registradas.

Catalogação cooperativa

A necessidade que o homem tem de ordenar e padronizar o mundo a sua volta, remonta a tempos antigos. Para se ter uma idéia, uma das listas de livros mais antiga de que se tem conhecimento data de 2000 antes de Cristo. A catalogação é uma das formas encontradas para organizar e ao mesmo tempo recuperar o conhecimento produzido.

A catalogação consiste em representar itens documentários de forma clara, sucinta e padronizada, individualizando-os e tornando-os acessíveis por várias formas de acesso para que o usuário possa recuperá-lo com rapidez e objetividade. A catalogação tem por objetivo a recuperação da informação, isto é, dizer onde está o documento.

Entende-se por item documentário qualquer objeto passível de ser catalogado como, livros, CDs, fotos, partituras, documentos eletrônicos, obras de arte e outros.

Os bibliotecários elaboram representações desses itens, de forma a simplificar a busca e recuperação. Essas representações abrangem tanto o aspecto físico dos itens como seu conteúdo. Através dessas representações criam-se diversos instrumentos como bibliografias, catálogos, boletins de serviço de alerta, entre outros (MEY, 1995).

Representar tem o mesmo significado de descrever, ou seja, ser a imagem ou a reprodução de algo, fazer a descrição fiel de um objeto ou um item documentário, tornando-o único e presente novamente. Segundo Marcondes (2001, p.64), "Etimologicamente, representar significa fazer alguma coisa presente novamente".

O fato de que o Homem pode manipular signos (símbolos que se referem a fenômenos naturais) na ausência dos próprios

fenômenos constitui uma diferença vital nas características dos animais (FOSKETT, 1980, p. 24).

A representação é uma atividade essencial e característica do ser humano. Para (MARCONDES, 2001, p. 63),

esta atividade mental e cognitiva foi mais útil ao homem em sua sobrevivência que seus atributos de força física. No amanhecer da humanidade, elas jogaram um papel fundamental na sobrevivência do homem.

A Catalogação Cooperativa pode ser definida como o trabalho de catalogação realizado de forma cooperativa, ou seja, trabalho realizado coletivamente por pessoas ou instituições que tenham objetivos em comum para o êxito de um mesmo propósito. Sendo assim, cada pessoa ou unidade de informação tem os seus direitos ao mesmo tempo que tem deveres a cumprir. As unidades de informação inseridas neste sistema têm participação ativa na confecção dos registros catalográficos e bibliográficos.

A representação *catalográfica* tem por objetivo a recuperação da informação, ou seja, dizer onde está o documento, diferentemente da representação *bibliográfica* que tem por objetivo disseminar a informação sobre o documento, ou seja, dizer que o item informacional existe. Essas diferenças podem ser observadas na Figura 1.

A pesquisa (recuperação de informações através de estratégias de busca) é uma atividade do bibliotecário, prontamente reconhecida pelo usuário. No entanto, é necessário que primeiro seja feito o tratamento e a organização da informação, com processos de construção de formas de representação como a catalogação, a classificação e a indexação para que dessa forma seja feito o seu correto armazenamento com o objetivo de possibilitar a recuperação e uso da informação.

Catalogação cooperativa automatizada

A cooperação sempre existiu na história humana, como uma alternativa de sobrevivência ou uma solução para os momentos de crise. Depois da Segunda Guerra Mundial, na década de 1940, houve uma produção espantosa de documentos, gerando um grande volume

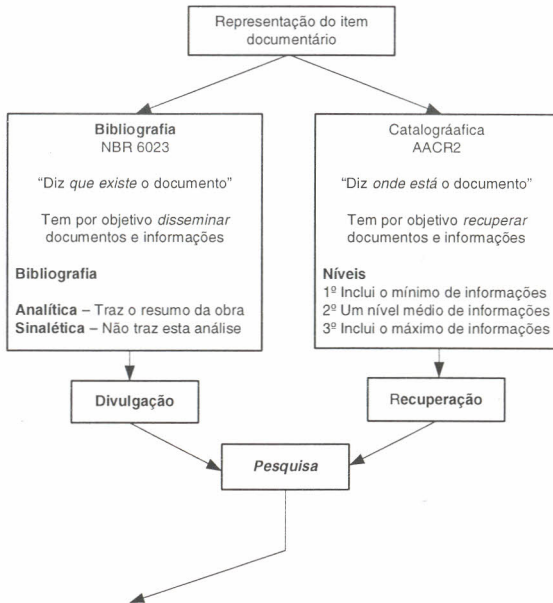


Figura 1 – Representação do item documentário.

de informação, que foi designado nessa época como “a explosão da informação”. Esta manteve-se, no entanto, parcialmente inacessível aos estudantes, cientistas e outros usuários, até o advento do computador e, conseqüentemente, da Internet.

A introdução e aplicação das tecnologias de informática para o processamento da informação bibliográfica constituíram-se em eventos importantes para a biblioteconomia nas últimas décadas. Atualmente bilhões de referências são armazenadas e tornadas disponíveis a cada ano, sob a forma digital.

No contexto histórico em que estamos vivendo, a era da Internet, a rede das redes e da informação, a ação cooperativa ganha enorme importância. “A Internet se parece cada vez mais com o sonho de Otlet (1934) de uma biblioteca universal” (OTLET, 1934 citado por MARCONDES, 2001, p.68). “E para que ela não evolua no sentido de um caos informacional, tem de ser trabalhada cooperativamente pelos profissionais de informação de todo o mundo” (MARCONDES, 2001, p.68). Hartung (1996, p.7) cita os valores do cooperativismo

que são: “ajuda mútua; responsabilidade; democracia; igualdade; equidade; solidariedade; honestidade; transparência; responsabilidade social e preocupação pelo seu semelhante”.

Para Souza (1999, p. 248),

A cooperação e o compartilhamento entre unidades de informação tem sido apresentados como a solução ideal para que possam atingir altos níveis de eficiência na busca e obtenção das informações necessárias a usuários que buscam subsídios para a realização de seus trabalhos, estudos, apoio ao ensino, ao desenvolvimento de pesquisas, à realização de negócios e outros. No que concerne à catalogação, as vantagens da cooperação são inúmeras e os bibliotecários, pelo menos na sua maioria, já se conscientizaram de que é uma grande perda de tempo, energia e dinheiro, duplicar a catalogação dos mesmos documentos adquiridos por diversas unidades de informação no país.

A cooperação e o compartilhamento de informações catalográficas e bibliográficas entre diversas instituições no ambiente automatizado somente se tornou possível com o desenvolvimento e adoção do Machine Readable Cataloguing – MARC. Este formato é considerado uma ferramenta importante que tem como objetivo conseguir um grau ótimo de normalização, padronização da representação descritiva automatizada dos dados informacionais, compatibilidade e transferência de dados legíveis por computador, permitindo a cooperação e o compartilhamento de recursos e serviços entre diferentes unidades de informação.

Formato de intercâmbio MARC 21

O Formato de Intercâmbio Machine Readable Cataloging – MARC ou seja, um registro catalográfico legível por computador, foi criado a partir da necessidade de minimizar esforços, reduzir custos dos processos técnicos e compartilhar informações. Como sabemos, toda invenção ou criação humana nasce da necessidade de facilitar a vida cotidiana e beneficiar o próprio ser humano, e com o formato MARC não foi diferente.

Devido à explosão de publicações científicas após a Segunda Guerra Mundial, buscava-se um meio de controlar, registrar e armazenar toda informação disponível. A década de 1960 é marcada pelo desenvolvimento dos recursos computacionais, e a Library of Congress – LC desenvolve e apresenta o formato MARC no período de 1965-1966. Isto conduziu a um projeto piloto, conhecido como MARC I, que investigou a praticabilidade de produzir dados de catálogos em formulários legíveis por computador.

Um trabalho similar estava em andamento no Reino Unido onde o Conselho de Bibliografia Nacional Britânica ajustou o projeto acima chamando-o de BNB MARC para examinar o uso de dados legíveis por computador na produção da Bibliografia Nacional Britânica impressa (BNB).

Em 1968, o projeto MARC II começou como um esforço anglo-americano para desenvolver um formato padrão de comunicação, ou seja, intercâmbio. Apesar da cooperação emergiram duas versões, UKMARC e USMARC.

O formato MARC torna-se base para outros formatos, a família cresce vertiginosamente e na década de 1970 contavam-se mais de cinquenta formatos. Listaremos aqui alguns exemplos: USMARC – Estados Unidos, UKMARC – Inglaterra, MONOCLE – França, CANMARC – Canadá, MARC/BR – Bélgica, IBERMARC – Espanha, MARC/MÉXICO – México, SAMARC – África do Sul, ANB MARC – Austrália, MAB1 – Alemanha, DN/MARC – Dinamarca, PICA – Holanda, NISSAT – Índia, MARCIS – Israel, ANNA MARC – Itália, JAPAN MARC – Japão, SWEMARC – Suécia, NORMARC – Noruega, CALCO – Brasil, e muitos outros.

Com tantos formatos, inevitavelmente surgiram diferenças e dificuldades no intercâmbio de informações. Para solucionar o problema de incompatibilidade a Library of Congress cria um formato internacional do MARC, o UNIMARC (MARC Universal).

Este formato cobriu originalmente monografias e séries, mas foi estendido em 1987 para incluir materiais não-livros. O formato foi mantido e desenvolvido sob a supervisão do Comitê permanente do UNIMARC.

Devido às razões econômicas da Biblioteca Britânica (UKMARC), da Biblioteca do Congresso Americano (USMARC) e da Biblioteca Nacional do Canadá (CAN/MARC) deu-se início ao movimento para

a harmonização desses três formatos MARC buscando formas de reduzir os custos da catalogação por meio da harmonização entre eles, para uma simplificação e padronização da catalogação, e dessa forma, uma biblioteca poderia utilizar os registros preparados por outras bibliotecas e assim reduzir custos nas instituições.

Depois de alguns estudos e tentativas, chegou-se à conclusão de que o formato UKMARC não poderia ser harmonizado completamente, em consequência das várias diferenças entre o UKMARC, o USMARC e o CAN/MARC. Alcançar a harmonização completa significaria que qualquer biblioteca e vendedores do sistema no Reino Unido teriam que fazer mudanças significativas ou os vendedores do sistema norte americano teriam que fazer mudanças caras no formato. Dessa forma, concordaram em excluir o UKMARC da harmonização dos formatos.

Iniciou-se então, em 1994, a harmonização dos dois formatos nacionais, o USMARC que foi desenvolvido na década de 1960 pela Library of Congress – LC (Biblioteca do Congresso Norte Americano), e o CAN/MARC desenvolvido em 1973 pela Biblioteca Nacional do Canadá, para controlar a sua catalogação legível por máquina.

Inicialmente foi utilizado o nome provisório de IMARC para o formato harmonizado, ou seja, MARC Internacional. Porém, durante a conferência da Federação Internacional de Associações e Instituições Bibliotecárias – IFLA em 1996, o nome IMARC causou uma reação muito negativa em vários países europeus, inclusive da IFLA, pois o mesmo foi visto como uma competição com o formato UNIMARC o qual fora patrocinado pela IFLA. Certamente não foi intenção do Canadá (CAN/MARC) competir com o UNIMARC. Assim, o IMARC estava descartado como um nome para os formatos harmonizados e depois de dois anos, buscando possíveis nomes, em outubro de 1998, a Library of Congress e o Escritório de padrões MARC propuseram o nome MARC 21 que significa MARC para o século 21, que foi aceito. MARC 21 passou a ser o nome dos formatos harmonizados do CAN/MARC e do USMARC. (STEWART, 1999).

O uso de um único Formato MARC padrão evita duplicação de trabalho e permite o melhor compartilhamento de recursos bibliográficos entre bibliotecas. A opção pelo uso do MARC 21 permite que as bibliotecas adquiram dados catalográficos previsíveis e confiáveis (FURRIE, 2000).

A estrutura de um registro MARC 21 é composta de três elementos básicos. São eles: O Líder, o Diretório e os Campos. Cada campo é associado a um número de 3 dígitos chamado “etiqueta”.

Observe na Figura 2 o exemplo de um campo. O número 245 é a etiqueta do campo que o define como o campo de título da obra.

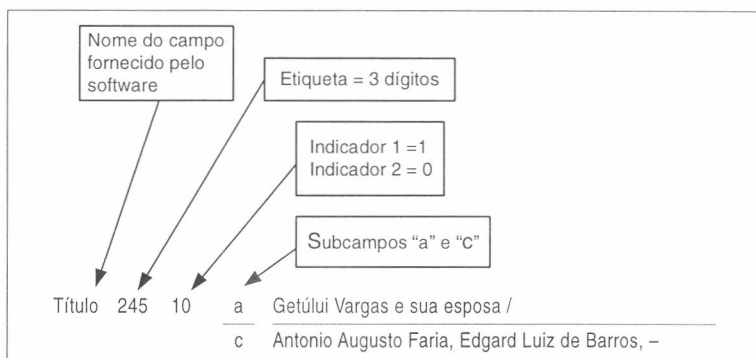


FIGURA 2 – Campo 245.

As etiquetas do MARC21 estão divididas em centenas; as divisões básicas são:

- 0XX Informação de controle, números, códigos
- 1XX Entradas principais
- 2XX Títulos, edição, imprensa (em geral, o título, menção de responsabilidade, edição e informação de publicação)
- 3XX Descrição física etc.
- 4XX Série
- 5XX Notas
- 6XX Campos de assunto
- 7XX Entradas secundárias
- 8XX Entradas secundárias de série etc.
- 9XX Livres para usos locais

A notação XX é freqüentemente usada para fazer referência a um grupo de etiquetas relacionadas. Por exemplo: 2XX refere-se a todas as etiquetas cujo primeiro dígito é “2” como, 210, 222, 240, 250 e assim por diante.

O Formato de Intercâmbio MARC21 se desdobra em três facetas. São elas: Formato de Entrada, Formato de Armazenamento Interno e Formato de Intercâmbio.

O Formato de Entrada é usado pelo catalogador na entrada dos dados. Para Balby (1995, p. 31),

um bom formato de entrada deve ser fácil de entender, ser próximo da forma de trabalho de quem coleta dados para o sistema, e possibilitar o preenchimento imediato, sem recurso a manuais separados de instruções.

O Formato de Armazenamento Interno diz respeito à alocação dos dados na memória do computador. Suas características mais desejáveis são eficiência e economia no armazenamento e recuperação dos dados (BALBY, 1995).

O Formato de Intercâmbio ou comunicação é gerado por programa, a partir do formato de armazenamento interno, com o fim de possibilitar a incorporação e interpretação de dados bibliográficos por sistemas e computadores diferentes. É voltado para a comunicação de dados bibliográficos entre computadores, e suas características mais desejáveis são a eficiência e a economia na troca de dados. Um formato de intercâmbio deve permitir a comunicação de dados entre computadores diferentes com o mínimo dispêndio em reformatação dos dados (BALBY, 1995).

No projeto do formato MARC, foi considerada a adoção de normas internacionais de descrição bibliográfica e de ferramentas de normalização.

O sistema adota, para a descrição bibliográfica, o Anglo-American Cataloguing Rules–2ª edição – AACR2, para a normalização utiliza a norma internacional de padronização ISO 2709 e a norma ANSI Z39.2, e para a recuperação e intercâmbio de dados catalográficos e bibliográficos em ambiente automatizado, o Protocolo Z39.50.

MARC 21: uma ferramenta, uma estrutura e um padrão para metadados

“Diferentemente da ISO 2709 e do protocolo Z39.50, o formato MARC 21 é uma ferramenta de domínio exclusivo do biblio-

tecário que o utiliza na atividade de descrição bibliográfica" (CÔRTE et. al., 1999, p.248).

Ferramenta é qualquer implemento ou objeto destinado a realizar uma operação ou executar um trabalho de qualquer natureza. Podemos dizer que qualquer que seja o tipo de trabalho realizado, todos nós usamos ferramentas. Por exemplo, o escritor cria com a mente mas necessita de uma ferramenta para anotar seus pensamentos. Um homem que dirige um caminhão, assim como uma pessoa que esteja utilizando um computador, estão fazendo uso de uma ferramenta, do mesmo modo que aqueles que os fabricaram. O emprego de ferramentas, o desenvolvimento da mecânica, a aplicação da eletricidade, foram os indispensáveis pioneiros de nossos recentes aperfeiçoamentos da automação, da aplicação de controles eletrônicos e das tecnologias de informática (ARNOLD e WHITE, 1965).

As ferramentas das novas tecnologias são as portas de entrada para o mundo globalizado que as unidades de informação de qualidade necessitam para estarem conectadas com essa nova realidade.

A qualidade de uma Unidade de Informação é representada pela sua capacidade de tratar as informações por meio dos processos de armazenamento, recuperação e disseminação com qualidade e padronização e a isso podemos chamar de otimização na gestão da informação. É com essa preocupação que "Bibliotecários e indexadores têm produzido e padronizado metadados por séculos" (MILSTEAD; FELDMAN, 1999, p.1, tradução nossa), e o Formato de Intercâmbio Machine Readable Cataloging – MARC 21, ou seja, um registro catalográfico legível por computador é considerado uma estrutura e um padrão para a construção de metadados produzidos na atualidade.

De acordo com as autoras Milstead e Feldman (1999), metadado não é nada de novo para os pesquisadores profissionais, pois eles têm sido capazes de melhorar a precisão de suas buscas através do uso de vocabulários controlados ou limitando as buscas aos descritores, identificadores, autores, títulos, assuntos e outros, por muitos anos. Embora metadado tenha-se tornado um *buzzword* no mercado de informação, o conceito é importante para autores e pesquisadores de informação eletrônica. Usado eficientemente, torna a informação acessível através de rótulos consistentes de seus conteúdos. Entende-se por *buzzword* um termo técnico que soa im-

portante, geralmente usado para impressionar pessoas leigas no assunto.

Para Milstead e Feldman, (1999, p.1, tradução nossa),

Como o homem que esteve escrevendo prosa toda sua vida sem saber, os bibliotecários e indexadores têm produzido e padronizado metadados há séculos. Ignorando este legado, uma imensa variedade de outros profissionais tem recentemente entrado neste campo, e muitos deles não têm idéia de que alguém mais já “esteve ali e feito isto” antes. Vários sistemas estão sendo desenvolvidos para diferentes e algumas vezes até para o mesmo tipo de informação, resultando numa atmosfera caótica de padrões.

Não importa o nome que se use, catalogação, indexação ou metadados; segundo Milstead e Feldman (1999, p.1, tradução nossa),

Seja qual for o nome que se use, catalogação, indexação ou metadados, o conceito é familiar para profissionais da informação. Agora o mundo eletrônico finalmente o descobriu. Até poucos anos atrás, somente alguns filósofos tinham ouvido a palavra “metadado”. Hoje, é difícil encontrar uma publicação que a ignore.

Metadados, como um novo conceito, surgiu da necessidade de ter representantes para o conteúdo dos bancos de dados que pudessem fornecer mais informações sobre eles, devido ao crescente número e tipos de objetos digitais disponíveis.

Para que os metadados alcancem os seus objetivos com sucesso é imprescindível o uso de listas ou bases de dados com assuntos e autoridades controladas, como as linguagens controladas ou tesouros. Por exemplo, se em um catálogo *online* coletivo não for adotado um padrão para o campo de responsabilidade intelectual, ou seja, autor, permitindo a entrada de várias formas como Silva, Maria; Silva, Maria A.; ou Maria Aparecida Silva, sem uma padronização para esse campo, caso o usuário queira reunir todas as obras dessa autora, ele provavelmente não conseguirá com as formas variantes da maneira como a indicação de autor foi efetuada nesse catálogo, ou seja, não foi usa-

da uma forma autorizada ou padronizada disponível em bases de dados de padronização oficial como a base BIBLIODATA ou Library of Congress – LC.

A padronização é fundamental tanto para a representação quanto para a recuperação da informação, sendo necessária a preocupação não somente com os campos para recuperação, mas também com a padronização do seu conteúdo.

Para El-Sherbini (2000, p. 188, tradução nossa), “Os Padrões para Metadados têm um importante papel no suporte ao uso de serviços e recursos eletrônicos. No entanto, para maximizar o uso destes e assegurar consistência, é necessário padrões universalmente aceitados que governem a criação de metadados”.

O Formato MARC é considerado a primeira estrutura e padrão desenvolvidos para a construção e representação de metadados. Possui um conjunto de elementos descritores complexos e rígidos para descrever informações catalográficas e bibliográficas, sendo necessário especialistas como o profissional bibliotecário para a correta descrição desses elementos (GRÁCIO; 2002; SOUZA; CATARINO; SANTOS, 1997).

Neste cenário surge a emergente Linguagem de Marcação Extendida – XML, possibilitando facilidades e interoperabilidades entre sistemas e melhor descrevendo as informações para intercâmbio de dados via Internet.

Linguagem de marcação XML

A eXtensible Markup Language – XML, ou seja, linguagem de marcação estendida, é descendente da Standard Generalized Markup Language – SGML, concebida especialmente para gerenciar, armazenar e transmitir dados via Internet. “Uma linguagem de marcação simplesmente define como você adiciona o significado aos dados, especialmente em um documento” (LIBERTY; KRALEY, 2001, p.6), por exemplo: <título> MARC 21 e XML </título>. Essas duas palavras entre tags, marcas ou marcação, são o título do documento. A XML “fornece pistas semânticas na forma de tags (marcas) sobre o significado dos dados” (LIBERTY ; KRALEY, 2001, p.5).

Segundo Ray (2001, p.2),

uma linguagem de marcação é um conjunto de símbolos que pode ser colocado no texto de um documento para demarcar e rotular as partes desse documento.

A linguagem eXtensible Markup Language (XML),

é o resultado do trabalho de um grupo de especialistas estabelecido em 1996 pelo W3C (World Wide Web Consortium),¹ com o objetivo de propor uma simplificação da SGML que fosse voltada às necessidades específicas da Web (BRYAN, 1998 citado por BAX, 2001, p.33).

"SGML é uma (meta)linguagem criada há aproximadamente 30 anos" [...] (EDWARDS, 1997 citado por BAX, 2001). É base de um novo paradigma, uma tentativa de padronização dos diversos formatos empregados para representar a informação. Foi reconhecida como um padrão ISO (8879) em 1986. As linguagens de marca ou marcação são aplicadas com vistas a um melhor gerenciamento da informação.

A linguagem XML ainda é recente; foi criada em 1996 e liberada oficialmente pelo World Wide Web Consortium – W3C (órgão responsável pelo seu desenvolvimento e manutenção) em 1998, dando-se, portanto, a sua introdução na Web no final da década de 1990, por isso ainda é muito discutida e analisada por especialistas no assunto. Liberty e Kraleý (2001, p.5) afirmam, "XML, uma das mais novas e sem dúvida, uma das mais poderosas tecnologias da Web".

A XML permite marcar os documentos levando em conta a semântica da informação, ou seja, cada pessoa ou instituição pode criar um conjunto de etiquetas que melhor represente os elementos que compõem os seus documentos (MARTÍNEZ GONZALEZ, 2000).

Explica Gazan (2000, p.14, tradução nossa):

A XML não é a HTML com novas *tags*. As *tags* ou etiquetas em um documento HTML são fixas, e fazem pouco mais que instruir um *browser* em como exibir o conteúdo. Elas não pro-

1 W3C – Organização que se encarrega do desenvolvimento e manutenção dos padrões da Web. Maiores informações visite o site www.w3c.org

vêm nenhuma ajuda nem posições seguras para as máquinas de busca que tentam ter acesso ao conteúdo de um documento. Por outro lado, as etiquetas XML descrevem o conteúdo do documento e expressam as relações hierárquicas entre elementos de dados, de modo muito semelhante ao dos campos dos registros em um banco de dados.

As duas linguagens de marcação HTML e XML, apesar de serem bastante parecidas, apresentam diferenças significativas. De acordo com Martínez González (2000); Castro (2001); Ramalho, (2002):

- A XML tem como finalidade descrever o conteúdo dos dados, enquanto a HTML se preocupa com a apresentação dos mesmos;
- A XML possui extensibilidade: as suas etiquetas são definidas de acordo com a necessidade de seus usuários, enquanto a HTML possui etiquetas predefinidas;
- A XML possui links mais potentes do que a HTML, permitindo recursos que esta não possui;
- A XML permite o armazenamento da informação e sua descrição em um mesmo local;
- A XML foi criada especialmente para a recuperação e o intercâmbio de dados;
- A sintaxe rígida da XML facilita o trabalho das aplicações informáticas, enquanto a sintaxe da HTML dificulta o tratamento dos seus dados pelos *browsers*, justamente pela falta de rigidez.

Observe os exemplos:

Em HTML:

```
<HTML>
<BODY>
Microcomputador Pentium IV, 1.5 GHz, 256MB de RAM, Monitor 17 polegadas,
Mouse, Teclado, estabilizador.
</BODY>
</HTML>
```

Figura 3 – Documento em HTML.

Fonte: FURGERI, 2001, p. 145.

Em XML:

```
<?xml version="1.0"?>
<MICROCOMPUTADOR>
  <MODELO> Pentium IV </MODELO>
  <VELOCIDADE>1,5GHz /VELOCIDADE>
  <RAM> 256MB </RAM>
  <MONITOR> 17 Polegadas </MONITOR>
  <TECLADO> Sim </TECLADO>
  <MOUSE> Sim </MOUSE>
  <ESTABILIZADOR>Sim </ESTABILIZADOR>
  <IMPRESSORA> Não </IMPRESSORA>
</MICROCOMPUTADOR>
```

Figura 4 – Documento em XML.

Fonte: FURGERI, 2001, p. 145.

Ao comparar os dois documentos, fica visível que a linguagem XML é bem parecida com a HTML, possuindo também *tags* para marcar as informações, no entanto existe uma diferença fundamental entre elas: a XML torna o documento mais “inteligente”, criando uma estrutura que demonstra claramente qual o significado das informações, enquanto a HTML se preocupa com a apresentação do conteúdo e não com o seu significado (FURGERI, 2001).

Quanto às ferramentas para criar documentos em XML, pode ser utilizado qualquer editor de texto ou processador de palavras, como o Notepad ou o Wordpad no Windows ou TeachText ou SimpleText no Macintosh. Por ser essa uma linguagem de código aberto assim como a HTML, podem utilizar diversos programas desenvolvidos por terceiros para facilitar a criação de arquivos e estruturas de dados (CASTRO, 2001; RAMALHO, 2002).

Muitos editores de HTML foram desenvolvidos para facilitar a criação de páginas HTML, mas atualmente as atenções das indústrias estão voltadas à criação de editores XML. Fazendo-se uma busca na Internet pode-se ter acesso e permissão para *download* de vários editores XML como XMLSPY, MSXML, PETER’S XML EDITOR, entre outros (RAMALHO, 2002).

Para Castro (2001), a XML em si é muito simples. São as tecnologias associadas a ela que ampliam seu potencial. São elas: Document Type Definition – DTD, XML Schema, eXtensible

Stylesheet Language – XSL, Xlink, Xpointer, Xpath, XML Namespaces entre outras.

Calcula-se que os números estatísticos do uso da XML atualmente utilizados pelas instituições estejam bem altos, devido a sua enorme aceitação.

Afirma Lam (2002, p.90, tradução nossa),

Desde a sua liberação oficial em 1998 pelo W3C, a XML foi rapidamente adotada e implementada por muitas indústrias. Nos últimos anos, tomaram-se muitas iniciativas e elaboraram-se projetos que fazem uso da XML para metadados de biblioteca (dados bibliográficos, controle de autoridades, cabeçalhos de assunto e outros).

Como podemos perceber, a XML hoje é não somente uma realidade mas uma necessidade. É a promessa de tornar a Internet mais “inteligente”, transformando o modo de armazenamento, recuperação e intercâmbio de informações e, principalmente, a economia da informação, com redução de custos operacionais (FURGERI, 2001) para as instituições, e desta forma causando grandes mudanças em diversas áreas. Torna-se, assim, necessário examinar o impacto que ela vem causando, inclusive no mundo bibliotecário.

Para Ortiz-Repiso Jiménez (1999, p.222, tradução nossa), “tanto [...] [XML] como MARC são linguagens de marcação descritiva que criam textos estruturados”. Mas a XML permite flexibilidades que não podem ser encontradas no MARC.

Estruturas de marcação MARC 21 e XML

O MARC, desenvolvido originalmente para descrever livros impressos, com o passar dos anos foi sofrendo adaptações para facilitar descrição a uma variedade de tipos de documentos, assim como acesso e localização da informação nos recursos da Web.

O MARC foi a ferramenta que possibilitou a concretização da Catalogação Cooperativa em ambiente digital e em rede, e serviu de base para outros padrões para metadados criados na atualidade (Government Information Locator Service – GILS; Dublin Core – DC; Content Standard for Digital Geospatial Metadata – CSDGM, entre outros).

Hoje o formato continua a ser usado como formato de intercâmbio e comunicação de dados e todos os grandes sistemas de catalogação suportam a importação e exportação de registros no formato MARC (FIANDER, 2001, p. 18, tradução nossa).

É uma família de formatos: bibliográfico, de autoridade, de cabeçalho de assunto, de classificação e informação de comunidade. Dá suporte, há mais de trinta anos, à comunidade bibliotecária e à indústria de software especializado.

A sua rica contribuição à padronização e à comunidade por redes é inquestionável. Porém, devido a sua estrutura rígida e complexa, projetada para atender as necessidades de uma época em que os catálogos impressos ou em microfimes eram a norma, tornou-se limitada para atender as diferentes necessidades informacionais de hoje (ROWLEY, 1994).

De acordo com Ortiz-Repiso Jiménez (1999, p.222, tradução nossa), “o formato MARC que consolidou as bases de cooperação e comunicação de informação bibliográfica no mundo bibliotecário tem, porém, uma série de características negativas” e limitações para atender as atuais demandas informacionais.

Ortiz-Repiso Jiménez (1999, p.222, tradução nossa) dá alguns exemplos, dentre os quais citamos:

- É trabalhoso, lento e custoso realizá-lo e devido a sua complexidade deve ser manipulado por profissionais qualificados.
- Adapta-se muito mal à informação estruturada hierarquicamente, já que se baseia numa estrutura plana. Foi projetado para descrever e permitir a informação de um registro bibliográfico e não para estabelecer relações entre registros.

Para Tennant (2002a, p.1, tradução nossa),

no MARC, os campos não são explicitamente rotulados, mas codificados com um esquema numérico que não pode ser lido por alguém não familiarizado com a sintaxe complexa.

Tennant coloca em pauta a questão da marginalização do formato, visto que na opinião do autor, “o MARC tem sido sempre um

padrão secreto, nenhuma outra profissão usa o MARC ou qualquer coisa parecida com ele”.

Segundo Fiander (2001, p.26, tradução nossa),

a rigidez e as irregularidades internas do MARC têm começado a causar problemas para catalogadores e usuários, e estão começando a ficar para trás em pesquisas atuais sobre padrões de descrições bibliográficas.

Mediante as características negativas e as limitações do MARC citadas anteriormente, de acordo com Ortiz-Repiso Jiménez (1999, p.222, tradução nossa), “[...] [a XML] proporciona uma qualidade de trabalho promissor por diversas razões”:

- Pode tratar informações hierarquicamente inter-relacionadas em tantos níveis quantos forem necessários.
- Não há limitações para os tamanhos dos documentos;
- É um padrão internacional adotado por um número crescente de instituições governamentais, de pesquisa e da indústria.

Miller citado por Johnson (2001, p.84, tradução nossa),

- Caracteriza a XML como o “formato de dados universal para a Web”, dizendo que é uma “SGML simplificada, que usa a sintaxe HTML, com facilidades da Web acrescentadas”.
- Emprega um conjunto único de caracteres, Unicode para representar todos os dados.

Miller, de acordo com Johnson (2001, p.84, tradução nossa), também observou que “um aspecto significativo da XML é a sua separação do conteúdo, sua apresentação e ligação, de forma que cada conteúdo possa ser controlado otimamente”.

Segundo Bax (2002), esta é a filosofia da XML (a separação da apresentação, do conteúdo e da ligação). O formatador na Figura 5 está representando a XSL, que possibilita diferentes apresentações de um mesmo documento XML. Observe:

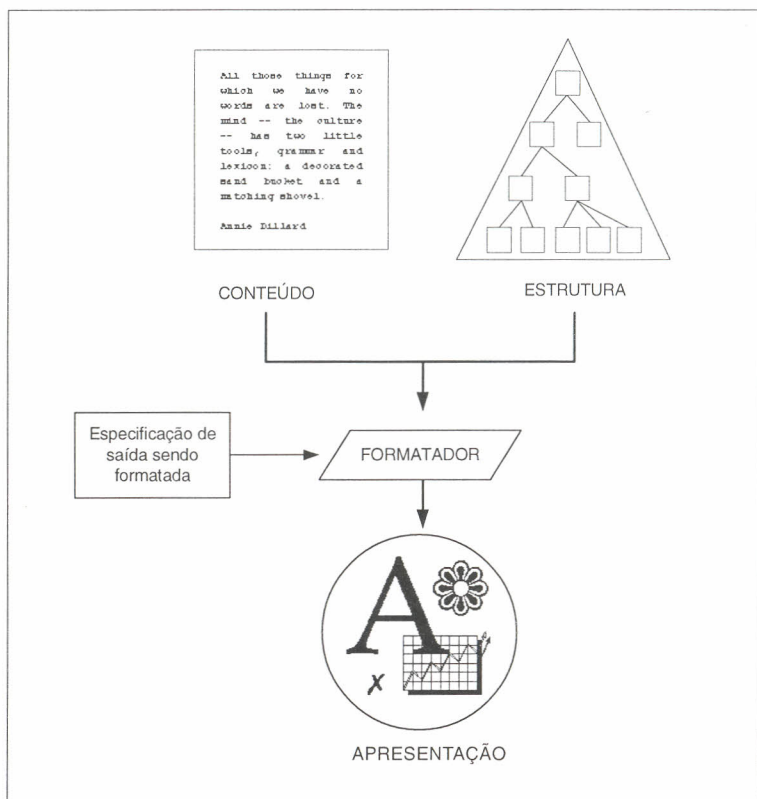


Figura 5 - Filosofia da XML.

Fonte: Bax, 2002, slide 18, tradução nossa.

Segundo Brun (2002, slide 3, tradução nossa), "o formato XML tem se apresentado como uma alternativa para representar e transferir registros MARC através da Internet e facilitar seu tratamento automatizado".

A união da estrutura de representações bibliográficas e catalográficas MARC e a linguagem XML serão extremamente valiosas ao considerarmos as diferenças entre as estruturas de marcação MARC e XML, conforme apresentou Bax (2002):

MARC	XML
<i>Estrutura Plana</i> O MARC é bidimensional (campo/ subcampo).	<i>Estrutura Hierárquica</i> A XML permite infinitas possibilidades de aninhamento.
<i>Formato proprietário</i> Restrições impostas à sua utilização.	<i>Formato aberto</i> Sem restrições. Qualquer pessoa ou instituição pode utilizá-la. Autonomia para a criação de DTDs, para necessidades específicas e marcações personalizadas.
<i>Dependência de plataforma</i> O MARC de uma certa maneira sempre foi proprietário, dependente de plataforma específica.	<i>Independência de plataforma</i> A XML é independente de plataforma, o software capaz de manipulá-la é o próprio navegador.
<i>Esquema de marcação</i> De difícil compreensão para pessoas não familiarizadas. Esquema codificado.	<i>Esquema de marcação</i> Compreensível por humanos e máquinas. Esquema semântico.
<i>Acesso restrito</i> Quase 100% restrito à comunidade bibliotecária.	<i>Acesso aberto</i> Enquanto que todos os outros profissionais da informação estão se aprofundando no uso da XML, além desta ser acessível a qualquer pessoa, qualquer área e a qualquer profissional.
<i>Registros tradicionais</i> Limitado aos registros tradicionais.	<i>Registros novos</i> A XML pode representar registros que não existiam: Textos longos (texto completo, biografia, resumo, sons – gráficos). Em XML encontramos os dados primários (o próprio documento) juntamente com dados secundários – (meta)dados.
<i>Sintaxe</i> Em geral, o MARC é uma sintaxe + um compromisso com uma implementação (um modelo de dados, limites de tamanhos específicos 99.999) Tendência ao desenvolvimento proprietário.	<i>Sintaxe</i> A XML é apenas uma sintaxe – separação mais clara entre a semântica da representação do registro e a semântica da aplicação.

Figura 6 – Esquema MARC e XML.

Porém, afirma Khurshid (2002, p.376, tradução nossa) :

Sugerir que o MARC deve ser substituído por esquemas de metadados não é aceitável por uma vasta maioria da comunidade bibliotecária. Ao invés disso, bibliotecários sugerem que deve ser permitida a coexistência do MARC e dos esquemas para metadados. A LC, OCLC e outras organizações já aceitaram a

importância disto e estão atualmente engajadas no desenvolvimento de mapas e caminhos cruzados, de forma que dados possam ser intercambiados.

No entanto, várias bibliotecas já iniciaram a migração e o uso do formato MARC em ambiente XML, como a Medlane,² com o formato XMLMARC,³ e a Library of Congress, com o MARCXML,⁴ entre outras.

Considerações finais

A XML, em relação à Biblioteconomia, poderia ser tida como uma inovação tecnológica para o MARC 21, vindo a ser considerada num futuro próximo como uma nova tecnologia, ou como o redesenho de uma série de tecnologias já existentes como SGML, HTML, MARC favorecendo a coexistência deste com outros esquemas para metadados, como o MARCXML e outros.

O período de transição de uma tecnologia para outra poderia ser breve se dependesse somente das tecnologias empregadas, mas geralmente esse período de transição é longo, pois não depende somente das tecnologias, mas também, e principalmente, da aceitação dessa nova tecnologia e de sua filosofia assim como da ruptura de paradigmas por parte dos profissionais de diversas áreas envolvidos na inovação tecnológica.

Em face deste novo espaço informacional, o profissional da informação se vê diante de uma nova direção na gestão da informação. Mas, como diz Alvarenga (2001, p.14), “mudam-se os meios, sofisticam-se os instrumentos e surgem nomes novos para designar coisas velhas. Entretanto, a essência das coisas permanece”.

E toda essa inovação tecnológica proporcionada pela XML e tecnologias associadas a ela vem dar condições para tornar realidade sonhos antigos, profetizados há séculos e ainda não concretizados de fato, como a grande biblioteca universal, o controle bibliográfico universal, entre outros.

2 Stanford University's Lane Medical Library

3 <http://xmlmarc.stanford.edu>

4 <http://www.loc.gov/standards/marcxml>

Através desta revisão de literatura observamos que o MARC foi o primeiro passo louvável para a consolidação da catalogação cooperativa em ambiente automatizado e, atualmente, ele é utilizado por bibliotecas espalhadas por todo o mundo, formando redes cooperativas de dados catalográficos e bibliográficos, que chegam a bilhões de registros disponíveis. No entanto, devido aos desenvolvimentos tecnológicos e ao intenso uso da Internet, torna-se necessário adaptá-lo a novas tecnologias emergentes, no caso a XML, o que já vem ocorrendo, porquanto que esta é uma linguagem de marcação atual e que atende as necessidades informacionais da Web.

Dessa forma, a utilização do formato MARC 21 juntamente com tecnologias associadas a ele, como a XML e outras, efetivaria uma catalogação cooperativa universal desde que haja uma política de adoção e investimentos nesse novo método e filosofia por parte dos profissionais e instituições.

Assim, o MARC 21 em ambiente XML permitirá a interoperabilidade entre diversas linguagens e estruturas de metadados, facilitando o intercâmbio de dados entre diversas instituições, independente do sistema ou plataforma utilizados, além de proporcionar uma melhor interação entre humanos e máquinas devido ao esquema semântico da XML.

O MARC 21, que até o momento representou eficientemente apenas registros catalográficos e bibliográficos, ou seja, registros tradicionais, em ambiente XML, será capaz de representar registros que o MARC 21 sozinho não suportaria, como textos completos, sons, gráficos e outros, representando tanto os dados primários (o próprio documento) como os dados secundários (metadados).

Podemos, através desta revisão de literatura, apontar desenvolvimentos tecnológicos que vêm ocorrendo na área da biblioteconomia e, em especial, na área da representação descritiva, como o XMLMARC desenvolvido pela Medlane e o MARCXML pela Library of Congress, entre outros.

Inclusive os grandes softwares de gerenciamento de bibliotecas utilizados por unidades de informação espalhadas pelo mundo inteiro já estão implementando uma camada XML em suas estruturas, como o ALEPH500, por exemplo, buscando, assim, facilitar o intercâmbio de dados entre as instituições por meio da interope-

rabilidade entre várias estruturas de metadados e sistemas heterogêneos utilizados (Informação verbal).⁵

Portanto, torna-se necessário que os profissionais da informação de hoje não só aceitem as mudanças, mas sejam agentes dessas, rompendo paradigmas, incorporando as mudanças, as inovações e as diversidades, visto que as necessidades sociais vigentes exigem isso.

Referências

- ALVARENGA, L. A teoria do conceito revisitada em conexão com ontologias e metadados no contexto das bibliotecas tradicionais e digitais. *Datagramazero – Revista de Ciência da Informação*, v2, n.6, dez/2001. Disponível em: <http://www.dgzero.org/dez01/Art_05.htm>. Acesso em: 29 jan. 2003.
- ARNOLD, P. ; WHITE, P. *A era da automação*. Rio de Janeiro: Lidador, 1965, 206 p. Tradução de: Hersílio Soares.
- BALBY, C. N. *Formatos de intercâmbio de registros bibliográficos: conceitos básicos*. Cadernos da F.F.C. Marília, v.4, n.1, 1995. p.29-35.
- BAX, M. P. Introdução às linguagens de marcas. *Ciência da Informação*, Brasília, v. 30, n.1, p.32-38, jan./abr. 2001.
- BAX, M. P. *XML – Introdução e impactos nas bibliotecas*. 2002. Slides. Disponível em: <www.bax.com.br>. Acesso em: 20 nov. 2002.
- BRUN, R. E. *XML en la descripción de recursos*. Sevilha, 2002. Disponível em: <<http://www.forpas.us.es/aula/xml/doc/09.XML%20en%20la%20descripci%C3%B3n%20de%20recursos%20MARC.ppt>>. Acesso em: 20 nov. 2002.
- CARVALHO, F. Ciência aguarda dilúvio de informação. *O Estado de São Paulo*, 25 mar. 1990, p.24.
- CASTRO, E. *XML para a World Wide Web*. Rio de Janeiro: Campus, 2001. 269 p. (Visual quickstart guide). Tradução de Hugo de Souza Melo.
- CÔRTE, A. R. et al. Automação de bibliotecas e centros de documentação: o processo de avaliação e seleção de softwares. *Ciência da Informação*, Brasília, v.28, n.3, p.241-256, set./dez. 1999.
- EL-SHERBINI, M. Metadata and the future of cataloging. *Library Computing*, [S.l.], v.19, n.3/4, p.180-191, 2000.
- FIANDER, D. Applying XML to the Bibliographic Description. *Cataloging & Classification Quarterly*, New York, v.33, n.2, p.17-28, 2001.

5 FONSECA, Marcelo Valle da. (Gerente da Ex Libris no Brasil). Palestra sobre: *Implantação de sistema de automação de biblioteca: Sistema ALEPH/2 – Metabuscadores de informação estruturada: Sistema Metalib* – realizada em 25 de novembro de 2003 no Anfiteatro II da Faculdade de Filosofia e Ciências – UNESP – Campus de Marília.

FOSKETT, D. J. *Informática*. In: Ciência da informação ou informática? Rio de Janeiro: Calunga, 1980. p.9-51.

FURGERI, S. *Ensino didático da linguagem XML*. São Paulo: Érica, 2001. 278 p.

FURRIE, B. *O MARC bibliográfico: um guia introdutório; catalogação legível por computador*. Brasília : Thesaurus, 2000. 95 p. Tradução de: Beatriz Valadares Cendón, Sonia Burnier, Maria Helena Santos e Natália Guiné de Mello Carvalho.

GAZAN, R. "The wonder years" of XML. *Library Computing*, [S.l.], v.19, n.1/2, p.13-17, 2000.

GRÁCIO, J. C. A. *Metadados para descrição de recursos da Internet: o padrão Dublin Core, aplicações e a questão da interoperabilidade*. 2002. 127 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2002.

HARTUNG, A. P. *O cooperativismo: ao alcance de todos*. 2.ed. Florianópolis: OCESC, 1996.

JOHNSON, B. C. XML and MARC: which is "right"? *Cataloging & Classification Quarterly*, New York, v.32, n.1, p. 81-89, 2001.

LAM, K. XML and global name access control. *OCLC Systems & Services*, [S. l.], v.18, n.2, p.88-96, 2002.

KHURSHID, Z. From MARC to MARC21 and beyond: some reflections on MARC and the Arabic language. *Library Hi tech*, [S. l.], v.20, n.3, p.370-377, 2002.

LIBERTY, J. ; KRALEY, M. *Aprendendo a desenvolver documentos XML para a Web*. São Paulo: Makron Books, 2001.274 p. Tradução de: Flávia Cruz.

MARCONDES, C. H. Representação e economia da informação. *Ciência da Informação*, Brasília, v.30, n.1, p.61-70, jan./abr. 2001.

MARTÍNEZ GONZÁLEZ, M. M. Extended Markup Language (XML) : Una solución para modelar documentos y sus interrelaciones basada en la semántica de la información. *Scire: Representación y Organización del Conocimiento*, Zaragoza, v.6, n.2, p.121-151, jul./dic. 2000.

MEY, E. S. A. Introdução à catalogação. Brasília: Brinquet de Lemos/Livros, 1995, 123 p.

MILSTEAD, J. ; FELDMAN, S. Metadata: Cataloging by Any Other Name. *Online*, Jan. 1999. Disponível em: <<http://www.onlineinc.com/online/OL1999/milstead1.html>>. Acesso em: 10 set. 2002.

NORONHA, D. P. ; FERREIRA, S. M. S. P. Revisões de Literatura. In: CAMPELLO, Bernadete Santos; CENDÓN, Beatriz Valadares; KREMER, Jeannette Marguerite (Org.). *Fontes de Informação para Pesquisadores e Profissionais*. Belo Horizonte, Ed. UFMG, 2000, p.191-198.

ORTIZ-REPISO JIMÉNEZ, V. Qué enseñamos después Del MARC? *Organización del Conocimiento en Sistemas de Información y Documentación*. Zaragoza, v.3, p.217-225, 1999. Actas del III Encuentro de ISKO-España Getafe, 19 al 21 de noviembre de 1997.

RAMALHO, J. A. *XML teoria e prática*. São Paulo: Berkeley Brasil, 2002, 146 p.
 RAY, E. T. *Aprendendo XML*. Rio de Janeiro: Campus, 2001. 372 p. Tradução de: Daniel Vieira.

ROWLEY, J. *Informática para bibliotecas*. Brasília: Brinquet de Lemos/Livros, 1994, 307 p. Tradução de: Antonio Agenor Brinquet de Lemos.

SOUZA, M. I. F. ; VENDRUSCULO, L. G. ; MELO, G. C. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. *Ciência da Informação*, Brasília, v.29, n.1, p.93-102, jan./ abr. 2000.

SOUZA, T. B. ; CATARINO, M. E. ; SANTOS, P. C. Metadados: catalogando dados na Internet. *Transinformação*, v.9, n.2, mai/ago, 1997.

SOUZA, T. B. Catalogação Cooperativa na Rede Bibliodata/CALCO: a repetitividade de registros bibliográficos no catálogo coletivo. In: SIMPÓSIO INTERNACIONAL DE BIBLIOTECONOMIA "DR. PAULO TARCÍSIO MAYRINK", 3., 1999, Marília. *Anais...* Marília: UNESP, 1999, p.247-258.

TENNANT, R. MARC Must Die. *Library Journal*, 2002a. Disponível em: <<http://libraryjournal.reviewsnews.com/index.asp?layout=article &articleid=CA250046&display=Digital+LibrariesNews&industry=Digital+Libraries&industryid=3760&verticalid=151>>. Acesso em: 20 set. 2002

STEWART, M. *MARC harmonization update: part I. background to MARC21*. In: CLA'99, 1999, Toronto. Disponível em: <<http://www.fis.utoronto.ca/people/affiliated/tsig/stewart.html>>. Acesso em: 18 dez. 2001.